

Article

Putative receptor-binding domain of bat-derived coronavirus HKU9 spike protein: evolution of betacoronavirus receptor binding motifs

Canping Huang, Jianxun Qi, Guang-Wen Lu, Qihui Wang, Yuan Yuan, Ying Wu, Yanfang Zhang, Jinghua Yan, and George F. Gao

Biochemistry, **Just Accepted Manuscript** • DOI: 10.1021/acs.biochem.6b00790 • Publication Date (Web): 03 Oct 2016

Downloaded from <http://pubs.acs.org> on October 10, 2016

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

Biochemistry is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Putative receptor-binding domain of bat-derived coronavirus HKU9 spike protein: evolution of betacoronavirus receptor binding motifs

AUTHOR NAMES

Canping Huang¹, Jianxun Qi², Guangwen Lu³, Qihui Wang⁴, Yuan Yuan⁵, Ying Wu², Yanfang Zhang^{2, 5}, Jinghua Yan⁴, George F Gao^{1,2,5,6,7*}

AUTHOR ADDRESS

¹National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing 102206, China

²CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

³West China Hospital Emergency Department (WCHED), State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, and Collaborative Innovation Center of Biotherapy, Chengdu, Sichuan 610041, China.

⁴CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China.

⁵School of Life Sciences, University of Science and Technology of China, Hefei, Anhui Province 230026, China

⁶Laboratory of Protein Engineering and Vaccines, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

⁷Research Network of Immunity and Health (RNIH), Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China

KEYWORDS

bat coronavirus HKU9; putative receptor binding domain (RBD); spike (S); structure; homologous inter-subdomain topology and binding mode; evolution

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 ABSTRACT

2 The suggested bat-origin for Middle East respiratory syndrome coronavirus (MERS-CoV) has
3 revitalized the studies on other bat-derived coronaviruses for the interspecies transmission
4 potential. Bat coronavirus (BatCoV) HKU9 is an important betacoronavirus (betaCoV) that is
5 phylogenetically affiliated with the same genus as MERS-CoV. The bat-surveillance data
6 indicated that BatCoV HKU9 has been widely spreading and circulating in bats. This highlights
7 the necessity of characterizing the virus for its potential of crossing species barriers. The receptor
8 binding domain (RBD) of the coronavirus spike (S) recognizes host receptors to mediate virus
9 entry and is therefore a key factor determining the viral tropism and transmission capacity. In
10 this study, the putative S RBD of BatCoV HKU9 (HKU9-RBD), which is homologous to other
11 betaCoV RBDs that have been structurally and functionally defined, was characterized via a
12 series of biophysical and crystallographic methods. By using surface plasmon resonance, we
13 demonstrated that HKU9-RBD binds to neither the SARS-CoV receptor ACE2 nor the MERS-
14 CoV receptor CD26. We further solved the atomic structure of HKU9-RBD, which is expectedly
15 compose of a core and an external subdomain. The core subdomain fold resembles those of other
16 betaCoV RBDs; whereas the external subdomain is structurally unique with a single helix,
17 explaining the inability of HKU9-RBD to react with either ACE2 or CD26. Via comparison of
18 the available RBD structures, we further proposed a homologous inter-subdomain binding mode
19 in betaCoV RBDs that anchors the external subdomain to the core subdomain. The revealed
20 RBD features would shed light on the betaCoV evolution route.

1 Introduction

Coronaviruses are large, enveloped and positive-stranded RNA viruses which can infect birds, animals and humans^{1, 2}. Taxonomically, these viruses are affiliated to the *Coronaviridae* family within the order of *Nidovirales*^{1, 3}. Ever since the 1930s when the first coronavirus of infectious bronchitis virus was isolated in chicken⁴, coronaviruses have expanded thus-far into four genera, *Alpha*, *Beta*, *Gamma*³ and *Deltacoronavirus*^{5, 6}, respectively. Of these, betacoronaviruses (betaCoVs) have drawn worldwide attention because of their pathogenic capacity and transmission potential to cause a global pandemic of human infections^{7, 8} and of their wide-spread and existence of enormous species in bats^{6, 9-11}. In 2002-2003, one representative betaCoV, the severe acute respiratory syndrome coronavirus (SARS-CoV) firstly emerged in China¹²⁻¹⁵ and then rapidly spread to other countries, leading to >8000 infection cases and >800 deaths⁷. In 2012, another betaCoV, named the Middle East respiratory syndrome coronavirus (MERS-CoV)¹⁶, was identified first in Saudi Arabia^{17, 18}. Despite the global efforts trying to control its transmission, MERS-CoV still spreads to affect multi-countries in the Middle East, Europe, North America and Asia, causing 1,800 confirmed infections and at least 640 deaths as of June 23th 2016 (based on the latest statistic data released by the World Health Organization⁸). Meanwhile, a human-infective betaCoV of HKU1 was isolated from a patient with respiratory disease in Hong Kong¹⁹. These unexpected outbreaks of betaCoV infection have posed a severe threat to the global public health and lead to enormous socioeconomic disruptions.

Phylogenetically, betaCoVs can be further categorized into four (A, B, C and D) evolutionary lineages/subgroups^{1, 3}. SARS-CoV is a typical lineage B member, while MERS-CoV is grouped in the C-lineage²⁰. Despite belonging to different subgroups, these two betaCoVs likely share similar interspecies transmission routes by “jumping” from its natural host/(s) to an intermediate

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 adaptive animal/(s) and finally to humans²¹. Current evidence clearly showed that SARS-CoV
2 originated from bats^{9, 22, 23} and possibly adapted in civets or raccoon dogs²⁴ before it infected
3 humans. Given the close phylogenetic relationship between MERS-CoV and a variety of bat-
4 derived coronaviruses (BatCoV) (e.g. HKU4, HKU5^{10, 25} and those recently identified in the
5 Middle East, Africa, Europe and Asia²⁶⁻³¹), it is widely accepted that the current MERS-epidemic
6 represents another bat-to-human transmission event related with a betaCoV, though its
7 intermediate host is shown, this time, to be dromedaries^{32, 33}. Notably, two recent studies
8 reported that BatCoV HKU4 could recognize human CD26, the MERS-CoV receptor³⁴, as a
9 functional entry receptor^{35, 36}, indicating its potential adaptation for human infection. These
10 continuously occurring yet unpredictable events of betaCoVs repeatedly crossing species barriers
11 highlight the pressing necessity of studies on other members of the genus for the characters
12 relevant to the interspecies transmission²¹.

13 The coronavirus spike (S) protein, which locates on the envelope surface of the virion, functions
14 to mediate receptor recognition and membrane fusion¹ and is therefore a key factor determining
15 the virus tropism for a specific species^{21, 37}. In most cases, coronaviral S will be further cleaved
16 into S1 and S2 subunits, and the capacity of receptor-binding is allocated to the S1 subunit¹. The
17 receptor binding domain (RBD) of betaCoV that directly engages the receptor is commonly
18 located in the C-terminal half of S1 (C terminal domain, CTD), such as in SARS-CoV³⁸, MERS-
19 CoV^{39, 40} and BatCoV HKU4³⁵, though in rare cases such as with mouse hepatitis virus (MHV)⁴¹,
20 the RBD region was identified in the S1 N-terminal domain (NTD). We previously characterized
21 structurally the MERS-CoV RBD (MERS-RBD) as a relatively independent entity composed of
22 a core and an external subdomain³⁹. The latter subdomain, which is topologically an insertion
23 between two scaffold strands of the core subdomain, presents a flat 4-stranded β -sheet surface to

1 contact the CD26 receptor³⁹. Similar topological arrangement of the core and external
2 subdomains into a structural unit for receptor engagement was also observed in SARS-CoV RBD
3 (SARS-RBD)³⁸. Nevertheless, SARS-RBD exhibits a unique loop-dominated external fold to
4 recognize human angiotensin converting enzyme 2 (ACE2)⁴² as a receptor. These observations
5 indicate that the homologous RBD regions of betaCoVs represent a key determinant in receptor
6 adaptation and cross-species transmission²¹.

7 BatCoV HKU9 is a representative betaCoV of the D-lineage¹¹. The virus was firstly identified in
8 bats in 2007 by next generation sequencing (NGS)¹¹. Though the isolation of live viruses has
9 been unsuccessful by far, its genomes are widespread in different bat species⁴³⁻⁴⁶. As people are
10 worrying its interspecies transmission potential, the features of its S protein, especially of the
11 homologous RBD region (HKU9-RBD), remain unavailable. This would be an indispensable
12 step towards understanding the pathogenesis of BatCoV HKU9. In addition, the atomic structure
13 of HKU9-RBD would provide requisite information in understanding the evolution of betaCoVs.
14 It is notable that MERS-RBD and SARS-RBD share a conserved core structure but differ in the
15 external fold to engage different receptors^{21, 38, 39}. Sequence features of betaCoV RBDs clearly
16 indicate that this scheme of subdomain-arrangement might be expanded to the whole
17 *betacoronavirus* genus, regardless the species. This notion was supported by our recent study on
18 BatCoV HKU4 RBD (HKU4-RBD) which exhibits a quite resembled structure to MERS-RBD³⁵.

19 In this study, we reported the structural and functional characterization on HKU9-RBD. The
20 solved structure expectedly contains a core subdomain homologous to those observed in other
21 betaCoV RBD structures and an external subdomain that is mainly α -helical. This unique
22 structural feature explains its inability to react with either human CD26 or ACE2, which is well
23 observed in our surface plasmon resonance (SPR) assay. Via comparison of available RBD

structures, we further showed that the detailed interactions, anchoring the external subdomain to the core subdomain, share similar patterns in betaCoV RBDs. We believe the observed core/external interacting mode represents another structural feature in the S that is reserved during the evolution of betaCoVs, in addition to the conservation in the fold for the core subdomain. Our study therefore further supports the notion that betaCoV S originates from the same ancestor and divergently evolves mainly in the RBD external region to engage variant receptors, thereby preparing for potential interspecies transmission.

Materials and Methods

Plasmid construction

The plasmids used for protein expression were individually constructed by insertion of the coding sequences for HKU9-RBD (S residues S355-N521, GenBank accession number: EF065513), MERS-RBD (S residues E367-Y606, accession number: JX869050), SARS-RBD (S residues R306-F527, accession number: NC_004718), human CD26 (residues S39-P766, accession number: NP_001926) and human ACE2 (residues S19-D615, accession number: BAJ21180) into the *EcoRI* and *XhoI* restriction sites of a previously modified pFastBac1 vector⁴⁷ which was engineered to include an N-terminal gp67 signal peptide coding sequence. For each protein, an engineered C-terminal hexa-his tag was utilized to facilitate protein purification. To prepare mouse IgG Fc fragment (mFc) fused proteins, the coding sequences of MERS-RBD, SARS-RBD and HKU9-RBD were fused with mFc sequence and then constructed into the pCAGGS vector respectively³⁵.

1 Protein expression and purification

2 The proteins used for crystallization and SPR analysis were prepared with the Bac-to-Bac
3 baculovirus expression system (Invitrogen) according to the manufacturer's instructions⁴⁸. In
4 brief, the verified pFastBac1 vector was transformed into the DH10Bac competent cells to
5 generate the recombinant bacmid. The bacmid was then extracted and transfected into Sf9 cells
6 to prepare the baculovirus stocks. Sf9 cells were further used to amplify the baculoviruses, while
7 High5 cells were used to express the protein.

8
9 The cell culture of High5 was collected 48 h post infection. Totally 4 L of cell culture of each
10 protein were collected and centrifuged at 6,500 rpm for 1.5 h to remove cell debris. After filtered
11 with 0.22 µm membrane, the supernatant was passed through two 5 ml HisTrap HP columns (GE
12 Healthcare) to capture the individual protein of interest. For MERS-RBD, SARS-RBD, human
13 CD26 and human ACE2, the bound proteins were detached from HisTrap with 20 mM, 50
14 mM and 300 mM imidazole in 20 mM Tris-HCl and 150 mM NaCl buffer (pH 8.0). After SDS-
15 PAGE determination, fractions detached with 300 mM imidazole were pooled and further
16 purified by a Superdex[®] 200 column (GE Healthcare). For HKU9-RBD, the bound proteins were
17 detached from HisTrap with 20 mM, 50 mM and 300 mM imidazole in 20 mM HEPES and 150
18 mM NaCl buffer (pH 7.0). Fractions detached with 50 mM and 300 mM imidazole were pooled
19 respectively and dialyzed overnight against 5 L 20 mM HEPES and 150 mM NaCl buffer (pH
20 7.0) to remove imidazole. The dialysates were concentrated and further purified by a Superdex[®]
21 200 column (GE Healthcare). Each protein was stored in the buffer that is used for purification.

To prepare mFc-fused proteins by mammalian cell expression system, the recombinant pCAGGS plasmids were confirmed with Sanger sequencing and then prepared with EndoFree Maxi Plasmid Kit (TIANGEN, Beijing). Each recombinant plasmid was transfected into 293T cells with 50 µg plasmid DNA per T75 plate using Polyethylimine (PEI, Polysciences Inc.). After 5 h incubation, the transfected cells were washed with PBS twice and then replaced with DMEM without serum. The cells were maintained for three days and the supernatant was harvested and replaced with fresh DMEM medium and then maintained for another four days. The harvested supernatants were pooled and concentrated and then mixed with two volumes 20 mM Trisodium phosphate (pH 7.0). The mixture was passed through a 5 ml HiTrap™ Protein A HP prepacked column (GE Healthcare) to capture the individual protein of interest. After remove of impure proteins with 20 mM Trisodium phosphate (pH 7.0), the bound protein was detached from the column with 100 mM glycine (pH 3.0). Each fraction was neutralized with 1 M Tris-HCl (pH 9.0). After SDS-PAGE determination, the detached fractions with interest protein were pooled and concentrated. The buffer of each protein was then changed to PBS (pH 7.0) for further experiments.

SPR assay

The BiAcore® experiments were carried out at 25 °C using a BiAcore® 3000 or BiAcore® T100 machines with CM5 chips (GE Healthcare). For all the measurements, an HBS-EP buffer consisting of 10 mM HEPES, pH7.4, 150 mM NaCl and 0.005% v/v Tween® 20 was used, and all proteins were exchanged into this buffer in advance. Firstly, the HKU9-RBD, MERS-RBD and SARS-RBD proteins expressed by insect cell were used for SPR assay using a BiAcore®

3000 machine. BSA (negative control), HKU9-RBD, MERS-RBD and SARS-RBD proteins were immobilized on the chip at about 1000 response units (RU), according to manufacturer's amine-coupling chemistry protocol (GE Healthcare). Gradient concentrations of human CD26 (0, 19.5 – 5000 nM) or human ACE2 (0, 39 - 625 nM) were then flowed over the chip surface. After each cycle, the sensor surface was regenerated via a short treatment with 10 mM NaOH. The equilibrium dissociation constants (binding affinity, K_D values) were analyzed using BIA evaluation[®] (BIAcore[®] software). In order to exclude the possibility that HKU9-RBD could be non-functional because of immobilization or missing some important post-translational modifications on the protein, we purified the mFc-fused HKU9-RBD proteins in mammalian cells and assembled the binding abilities to CD26 or ACE2 proteins using a captured SPR method by a BIAcore[®] T100 system. The CM5 chip was immobilized with anti-mouse antibody for flow cell (FC) 1 and 2. The mFc-fused RBD proteins were then injected and captured on the FC 2, while FC 1 was used as negative control. Human CD26 or human ACE2 proteins were then injected and the binding responses were measured. The immobilized anti-mouse antibody was regenerated with 10 mM Glycine, pH 1.7 (GE Healthcare)..

Crystallization

The crystallization trials were performed with 1 μ L protein mixing with 1 μ L reservoir solution and then equilibrating against 100 μ L reservoir solution at 4°C by the vapor-diffusion sitting-drop method. The initial crystallization was screened using the commercially-available kits. Diffractive crystals of the HKU9-RBD protein were finally obtained in a condition consisting of 0.1 M sodium citrate tribasic dihydrate pH 7.0 and 12% PEG 20,000 with a protein concentration

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

of 2.2 mg/mL. Derivative crystals were obtained by soaking the crystals in the reservoir solution containing 1mM KAuBr₄•2H₂O for 48 h at 4 °C.

Data collection, integration and structure determination

For data collection, all crystals were flash-cooled in liquid nitrogen after a brief soaking in reservoir solution with the addition of 20% v/v glycerol. The diffraction data for the native (wavelength: 1.03906 Å) and Au derivative crystals (wavelength: 1.03906 Å) of HKU9-RBD were collected at Shanghai Synchrotron Radiation Facility (SSRF) BL17U. All data were processed with HKL2000⁴⁹. The ice-rings that form in the crystal flash cooling process were excluded in data processing and the final overall completeness for the data set is 97.1%.

The structure of HKU9-RBD was determined by SAD method. After location of Au sites by SHELXD⁵⁰ with the Au-SAD data, the identified positions were then refined and the phases were calculated with SAD experimental phasing module of PHASER⁵¹. The real space constraints were further applied to the electron density map in DM⁵². The initial model was built with Autobuild in the PHENIX package⁵³. Additional missing residues were added manually in COOT⁵⁴. The final model was refined with phenix.refine in PHENIX⁵³ with energy minimization, isotropic ADP refinement, and bulk solvent modeling. The stereochemical qualities of the final model were assessed with MolProbity⁵⁵. The Ramachandran plot distribution for the residues in the HKU9-RBD structure were 94.64, 5.36 and 0% for favored, allowed and outliers regions, respectively. Data collection and refinement statistics are

summarized in Table 1. All structural Figures were generated using PyMol (<http://www.pymol.org>).

Results

HKU9-RBD does not bind either the SARS-CoV or the MERS-CoV receptor

We first characterized the sequence of BatCoV HKU9 S by using a series of bioinformatic methods. This 1274-residue protein exhibits typical features of coronavirus S proteins (e.g. the presence of characteristic heptad repeats 1 and 2 in the S2 subunit), though the S1/S2 cleavage site potentially processed by furin-like proteases was not detected (Fig. 1A). Along the full-length protein, the amino acid sequence identity between BatCoV HKU9 S and other betaCoV Ss is rather limited (e.g. 27.9% to MERS-CoV S, 28.0% to HKU4 S and 30.4% to SARS-CoV S). Nevertheless, we were able to identify the RBD region based on the characteristic cysteine residues of the core subdomain (Fig. 1B), which were shown, in the thus-far available RBD structures^{35, 38, 39, 56}, to form three conserved disulphide bonds stabilizing the core fold. The subsequent HKU9-RBD was allocated to the S region spanning residues 355-521 (Fig. 1A). In comparison to other RBD sequences, HKU9-RBD exhibits a comparable length in the core subdomain (Fig. 1B) but is dramatically shortened in the external region (Fig. 1C).

To test if HKU9-RBD could react with either the SARS-CoV receptor ACE2⁴² or the MERS-CoV receptor CD26³⁴, the RBD and the receptor-ectodomain proteins were individually prepared in insect cells and purified to homogeneity. The ligand/receptor interaction was then characterized via SPR Biacore® by flowing through ACE2 or CD26 over the immobilized RBD proteins. As expected, potent interactions were observed for both the SARS-RBD/ACE2 (K_D

value: 0.265 μ M) (Fig. 2A) and the MERS-RBD/CD26 (K_D value: 52.8 nM) (Fig. 2B) binding pairs. The revealed kinetics were very similar to those reported in the previous studies^{35, 39}, validating the integrity of our testing system. Under the same condition, however, neither ACE2 (Fig. 2C) nor CD26 (Fig. 2D) interacted with HKU9-RBD. In order to exclude the possibility that HKU9-RBD could be non-functional because of immobilization or missing some important post-translational modifications on the protein, we purified the mFc-fused HKU9-RBD proteins in mammalian (293T) cells and assembled the binding abilities to CD26 or ACE2 proteins using a captured SPR method. In the same way, there was no detectable binding of mFc-fused HKU9-RBD to ACE2 or CD26 (Fig. 1G), while the mFc-fused SARS-RBD protein bound to ACE2 (Fig. 1E) and the mFc-fused MERS-RBD bound to CD26 (Fig. 1F) well. BatCoV HKU9, therefore, could utilize neither the SARS-CoV receptor nor the MERS-CoV receptor for cell entry. Rather, it must utilize a unique cellular receptor for entry.

Crystal structure of HKU9-RBD

We further set out to investigate the structural features of HKU9-RBD via crystallography. The protein was successfully crystallized, and a dataset of 2.1 angstrom ($\text{\AA}$) was collected (Table 1) and the structure was solved by using single-wavelength anomalous diffraction (SAD) method. The solved structure, with an $R_{\text{work}} = 0.1700$ and an $R_{\text{free}} = 0.2006$, contains a single molecule in the crystallographic asymmetric unit. Clear electron densities were traced for 176 consecutive HKU9-RBD residues, extending from S355 to A520. These amino acids fold into a compact structure, which can be further divided into two subdomains as schemed in the other RBD structures^{35, 38, 39}. The core subdomain comprises 8 β -strands and 6 helices (α or 3_{10}). Five long

1 strands ($\beta c1$ - $\beta c5$) were arranged in an anti-parallel manner, forming the scaffold center of the
2 core (core-center). This core-center sheet was further wrapped by the surface helices and loops.
3 It is notable that the 6 helices (H1-H6) were sporadically distributed on the two sheet faces,
4 thereby leading to an overall globular fold for the core subdomain. On one lateral side of the
5 core-center sheet, the external subdomain covers the core like a hat; while on the distal opposite
6 side, three small strands ($\beta p1$ - $\beta p3$) constitute a parallel peripheral sheet (core-peripheral),
7 clinching the N- and C-termini of HKU9-RBD in the close proximity. As expected, the
8 characteristic cysteine residues (Fig. 1B) form three disulphide bonds in the core subdomain,
9 further stabilizing the core structure from the interior. Of these, two (C357/C381 and
10 C411/C517) locate in core-peripheral, contributing to the RBD-termini orientation; one
11 (C399/C452) resides in core-center, linking strands $\beta c2$ and $\beta c4$ (Fig. 3). Overall, the residue
12 boundaries of the core subdomain observed in the structure are quite consistent with those
13 deduced from the sequence alignment result (Fig. 1B).

14 The external subdomain of HKU9-RBD consists of 42 residues from L458 to V499 (Fig. 1C).
15 These amino acids extend out of strand $\beta c4$ of the core-center sheet, first orient as a loop along
16 the core subdomain like a clamp, then fold back to form a solvent exposed α -helix (H1') and
17 finally proceed into the core strand $\beta c5$ (Fig. 3). This observed structure differs dramatically
18 from those of SARS-RBD and MERS-RBD, which are shown to be devoid of any helical
19 components in the external region^{38, 39}. The unique external fold of HKU9-RBD could well
20 explain its inability to bind either ACE2 or CD26.

22 Structural conservation of the RBD core subdomain in betaCoVs

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Previously, three betaCoV RBD structures have been reported, including one lineage B structure (SARS-RBD³⁸) and two lineage C structures (MERS-RBD³⁹ and HKU4-RBD³⁵). These structures indicated an interspecies conservation in the core fold among betaCoVs³⁹. BatCoV HKU9 is a representative member of the betaCoV D-lineage¹¹. We therefore compared the currently available RBD structures with the HKU9-RBD structure solved in this study. As expected, a significant similarity was observed in the core subdomain (Figs. 4A-D). Superimposition of the core-structures revealed the root-mean-square deviation (rmsd) values ranging from 0.66 to 2.82 Å (Table 2), demonstrating the quite resembled core fold (though low sequence identity) among the four RBDs. The most conserved part was seen in the core-center sheet. This 5-stranded scaffold element as well as the single inter-strand disulphide bond were invariably reserved in all the structures. In the core-peripheral region, however, small variance in the strand composition was noted. In HKU9-RBD, it contains three short β-strands, arranging into a small parallel β-sheet. Both SARS-RBD and MERS-RBD retained two of these strands, whereas HKU4-RBD is devoid of any detectable strand elements in this region. Despite the observed difference in strand formula, the core-peripherals of the RBDs exhibit similar orientation and present the same scheme of arranging the domain N-terminus into close proximity to its C-terminus. Extra common features in core-peripheral lie in the two disulphide bonds in the region, which were structurally and topologically conserved in the four structures (Figs. 4A-D).

In contrast to the core-conservation, the external subdomains of the four RBDs were divergent in structures. HKU9-RBD presents a single H1' helix in the external region, whereas SARS-RBD is loop-dominated, but contains two extra small β-strands. The external subdomains of MERS-RBD and HKU4-RBD, however, resemble each other and are predominantly a rigid β-sheet

1 composed of four β -strands. Despite the structural irrelevance, the external subdomains are
2 clearly topological equivalents in these structures, being present as an insertion between two
3 core-center strands (Figs. 4A-D).

4 5 **Homologous interaction mode anchoring the external subdomain to the core subdomain**

6 By superimposition of the available RBD structures, we unexpectedly identified two major
7 elements in the external subdomain that could be well-aligned (Fig. 5A). The first element
8 (element 1) spans about 7 residues (Y464-F470 in HKU9-RBD, Y438-R444 in SARS-RBD,
9 Y497-C503 in MERS-RBD and Y501-C507 in HKU4-RBD) (Figs. 1C&5C) and proceeds along
10 the core subdomain surface to be lodged between the H2 and H6 helices (based on the secondary
11 element definition of HKU9-RBD) (Fig. 5A). The second element (element 2) contains 8 amino
12 acids (P471-Q478 in HKU9-RBD, K447-D454 in SARS-RBD, L517-S524 in MERS-RBD and
13 Y522-S529 in HKU4-RBD) (Figs. 1C&5C), extending as a curved loop covering helix H6 of the
14 core subdomain (Fig. 5A). It is interesting that these two elements were “saddled” upon the core-
15 helices, anchoring the external subdomain to the core subdomain (Fig. 5A). We therefore further
16 explored the amino acid interaction details at this core/external interface.

17 Each element residue was scrutinized for both the side-chain orientation and the inter-subdomain
18 interactions. To facilitate the analyses and comparison, the two elements were assigned a
19 position marker for each amino acid (a-g for element 1 and a-h for element 2) (Figs. 5B&5C). In
20 element 1, the a-residue is invariably a tyrosine in the four RBDs. This amino acid oriented its
21 bulky side-chain towards the core subdomain, providing strong hydrophobic contacts. An extra
22 side-chain H-bond was also observed at this position in HKU9-RBD. The b-residue extended

1 away from the core surface and exhibited little conservation. The residue, however, invariably
2 contributed to the subdomain anchoring by providing a main-chain H-bond. Following b, the
3 amino acids were preferably facing towards the core at position c and spreading parallel to the
4 core surface at position d. Multiple van der Waals (vdw) contacts and conserved main-chain H-
5 bonds were observed at these two positions, respectively. Small discrepancy was seen in SARS-
6 RBD, which orients its c-residue outwards for the bulky solvent region. The remnant three
7 element 1 amino acids at positions e, f and g were distant from the core subdomain and
8 completely solvent exposed, therefore contributing little to the core/external interactions (Figs.
9 5B&5C).

10 In element 2, both the a- and d-residue were oriented parallel to the surface of the core
11 subdomain. The configuration allows the amino acids to provide apolar vdw contacts to
12 strengthen the core/external subdomain-binding. A certain extent of amino acid conservation was
13 observed at position d where a proline is favored to facilitate the turning of the loop. Following
14 these two positions, the b- and e-residue inserted their side-chains into two surface pockets of the
15 core subdomain. At position b, the residue is conservatively of hydrophobic and middle-sized
16 side-chain (Val/Ile/Leu). It is accommodated in a shallow apolar pocket, creating strong stacking
17 forces bonding the core and external subdomains. In addition, the residue also contributes a
18 main-chain H-bond to the subdomain binding. For the e-residue, its accommodating pocket is
19 deep and of large size, therefore allowing for amino acid variance at the position (Gly in HKU9-
20 and HKU4-RBD, Phe in SARS-RBD and Asn in MERS-RBD). In the four RBDs, this e-residue
21 invariably H-bonds with the core subdomain residue via the main-chain atom, but may optionally
22 provide the side-chain H-bond interactions (e.g. in MERS-RBD) or the multi-vdw contacts (e.g.
23 in SARS-RBD). Extra core/external interactions in this region were further observed at position

g, where the residue is oriented either in parallel to or towards the core subdomain and thereby contributes to the binding via hydrophobic and side-chain H-bond interactions. It is also of interest that these important interface residues of element 2 are regularly interspersed by amino acids at positions c, f and h, which are solvent exposed and rarely interact with the core subdomain (Figs. 5B&5C).

In summary, the four coronaviral RBD structures presented homologous amino acid interaction patterns for the inter-subdomain binding. The binding relies mainly on two elements in the external subdomain, which are oriented similarly in these structures (Fig. 5A). Despite the less conservation in the element sequences, the side-chain orientation and the interaction modes (hydrophobic, vdw or H-bond contacts) at each position are, in most cases, similar or homologous (Fig. 5C).

Discussion

Bats have been implicated harboring the largest natural genetic pools for new coronaviruses or coronaviral genes. A majority of the betaCoVs could be traced, for the origin, back to bats⁵⁷, *e.g.* a recent study isolated, in Chinese horseshoe bats, a live SARS-like coronavirus that can utilize the SARS-CoV receptor of ACE2 for cell entry²², thereby providing the strongest evidence for the bat-origin of this pandemic human pathogen. In addition, two studies reported the identification of gene fragments in bats that are almost identical to those of MERS-CoV^{26, 29}, indicating that MERS-CoV likely also originates from bats. Noted the recent reports showing the adaptation of batCoV HKU4 for binding to human cells via recognizing CD26³⁵, it is an urgent issue to prepare for the unforeseeable events of potential interspecies transmission by other bat-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

derived betaCoVs. BatCoV HKU9 is an important lineage D betaCoV¹¹ and has been demonstrated to be widespread and be circulating in different bat species⁴³⁻⁴⁶. Noted the determinative role of the coronaviral S RBD in the process of crossing species barriers (as has been structurally illustrated in other coronaviruses^{35, 38, 39}), we characterized the homologous RBD protein of batCoV HKU9 S for the structural and functional features. The solved structure revealed a core subdomain that resembles those observed in SARS-, MERS- and HKU4-RBD but a unique external subdomain that is composed of a single helix. Since the RBD external subdomain contains the key motifs (denoted receptor binding motif, RBM), interfacing with the receptor^{35, 38, 39}, the unique external fold of HKU9-RBD therefore well accords with our functional data showing its inability to react with either ACE2 or CD26. It remains to be investigated which host molecule could be recognized by HKU9-RBD as a functional cell entry receptor. Nevertheless, taking into account the single helical component in the external subdomain of HKU9-RBD, the RBM is expected to locate on the solvent-access side of the helix, which might facilitate future attempts for receptor identification.

It should be noted that coronavirus RBDs are not necessarily locating in the C-terminal half of the S1 subunit. Previous structural and mutagenesis data showed that the RBD of MHV S locates in the S1 N-terminal half (NTD)^{41, 58}. Nevertheless, the current data seem to favor a notion that CTD prioritizes over NTD to function as the receptor binding entity, as the majority of the coronaviruses (e.g. SARS-CoV³⁸, MERS-CoV³⁹, batCoV HKU4³⁵, human coronavirus NL63⁵⁹ and transmissible gastroenteritis virus⁶⁰ etc) harbor a CTD as RBD. It is interesting that the current available betaCoV CTD/RBD structures^{35, 38, 39} all clinch the N- and C-terminus on the same side opposite to where the external subdomain locates. This arrangement mode would lead the S1 N-terminal half to being sterically underneath the C-terminal half, thereby projecting the

CTD distant from the viral envelope for a trans-interaction with the receptors. Our structural study demonstrated that HKU9-RBD reserved the same character and therefore stands a better chance of being the authentic receptor binding entity.

In comparison to SARS-, MERS-, and HKU4-RBD whose structures are available^{35, 38, 39}, HKU9-RBD differs in the external fold but reserves a resembled core subdomain structure. The most conserved part lies in the core-center sheet which is composed of five anti-parallel strands and functions as the scaffold of the core subdomain. The sheet is sterically and structurally conserved in all the RBD structures. Additional conserved elements include the core-center helices and core-peripheral. Nevertheless, these elements could vary in the secondary-element compositions. E.g. A recent study on the structure of HKU4-RBD³⁵ showed that its N-terminal-most part does not fold into a characteristic helix as observed in the structures of SARS-³⁸ and MERS-RBD³⁹. For core-peripheral, the amount of strands was found to be varying from zero (as in HKU4-RBD) to three (as in HKU9-RBD) (Fig. 4). This has added dramatic complexities to the nomenclature of the RBD secondary elements. The situation would be even worse when taking into account of the external subdomains which could vary significantly in structure. We therefore suggest to designate the core-center strands and helices as β cs (β c1- β c5) and Hs (H1, H2...etc), respectively, and to refer to the core-peripheral strands as β ps (β p1, β p2...etc) and the external elements as H's (H1', H2'...etc) or β 's (β 1', β 2'...etc). This terminological strategy should be able to facilitate the comparison of homologous RBD structures and to reflect the fact that the external subdomain is topologically an insertion between two equivalent core-center strands.

The long evolutionary history, high mutation rates and many genetic artifices of RNA viruses often lead to conundrums in the virus origin study. It would be even more difficult to track the

1 evolutionary traces in the viral surface proteins which are normally under great evolutionary
2 pressure. The evolutionary records, however, are more likely to be conserved in the tertiary
3 structures than in the amino acid sequences. In betaCoVs, the inter-lineage sequence identity in
4 the S RBD is rather limited. Nevertheless, we observed several conserved features in the
5 betaCoV RBD structures. These include: 1) a conserved core-center as the scaffold of the core
6 subdomain; 2) a resembled core-peripheral where the RBD termini are clinched in proximity; 3)
7 a similar topological arrangement of the external subdomain as an insertion between two core-
8 strands; and 4) a homologous inter-subdomain binding mode anchoring the external subdomain
9 to the core subdomain. The features indicate a common ancestor S protein which divergently
10 evolves into different species. During evolution, the core subdomain is structurally reserved,
11 whereas the external subdomain folds into variant structures to engage different receptors. It is
12 also noteworthy that the aforementioned core features have been structurally validated in
13 betaCoV lineages B (SARS-RBD), C (MERS- and HKU4-RBD) and D (HKU9-RBD) but not
14 yet in lineage A. Structural studies on the equivalent S RBDs of the lineage A members should
15 be conducted in the future.

16
17 FIGURES
18 **Figure 1. Sequence features of HKU9-RBD.** (A) A schematic representation of the BatCoV
19 HKU9 S. The indicated domain elements were defined based either on the pairwise sequence
20 alignment results or on the bioinformatics predictions. The signal peptide (SP), transmembrane
21 domain (TM) and heptad repeats 1 and 2 (HR1 and HR2) were predicted with SignalP 4.0 server,
22 TMHMM server and Learncoil-VMF program, respectively; while the N-terminal domain

(NTD) and RBD were deduced by aligning with the N-terminal galectin-like domain of murine hepatitis virus S and with MERS-RBD, respectively. The S1/S2 site potentially cleaved by furin-like proteases could not be ascertained and is therefore labeled with a question mark. (B, C) Structure-based alignment of the HKU9-, SARS-, MERS- and HKU4-RBD sequences. The arrows and spiral lines indicate strands and helices, respectively. These secondary structure elements were labeled as illustrated in Fig. 3. The conserved cysteine residues that form three disulphide bonds in the structures are marked with Arabic numbers 1-3. The core subdomain is conserved among the four RBD structures; but the external subdomain is structurally irrelevant. We therefore presented the sequences separately. The two elements that anchor the external subdomain to the core subdomain were highlighted with the black boxes. (B) Core subdomain sequence. (C) External subdomain sequence.

Figure 2. Characterization of HKU9-RBD by SPR assays. The indicated RBD proteins expressed by insect cells were immobilized on CM5 chips and tested for the binding with gradient concentrations of human ACE2 or CD26 using a BIAcore® 3000 machine. The recorded kinetic profiles are shown. (A) Human ACE2 to SARS-RBD. (B) Human CD26 to MERS-RBD. (C) Human ACE2 to HKU9-RBD. (D) Human CD26 to HKU9-RBD. Clearly shown is that HKU9-RBD does not bind either ACE2 or CD26, in the context of SARS-RBD and MERS-RBD bind their respective receptors. Then we purified the mFc-fused HKU9-RBD proteins in mammalian (293T) cells and assembled the binding abilities to CD26 or ACE2 proteins using a captured SPR method by a BIAcore® T100 system. The anti-mouse antibodies were immobilized on CM5 chips. The mFc fused RBD proteins were then captured (3 µg/mL for 60 seconds) by the antibodies, and tested for the binding with human ACE2 or CD26. (E) The mFc fused SARS-

1
2
3 1 RBD (SARS-RBD-mFc) did not bind to CD26, while bound to ACE2 well. (F) The mFc fused
4
5 2 MERS-RBD (MERS-RBD-mFc) did not bind to ACE2, while bound to CD26 well. (G) The mFc
6
7 3 fused HKU9-RBD (HKU9-RBD-mFc) does not bind either ACE2 or CD26.
8
9
10
11 4
12
13

14 5 **Figure 3. Crystal structure of HKU9-RBD.** The core and external subdomains are colored
15
16 6 magenta and green, respectively. The core subdomain is further divided into a center region
17
18 7 (core-center) and a peripheral region (core-peripheral), which are encircled for highlight. The
19
20 8 core-center strands and helices are labeled $\beta c1-\beta c5$ and H1-H6, respectively; while the core-
21
22 9 peripheral strands are marked $\beta p1-\beta p3$. The disulphide bonds and the RBD termini are labeled.
23
24 10 The core subdomain is further presented in a surface representation in the right panel to highlight
25
26 11 the top positioning of the external subdomain like a hat.
27
28
29
30
31
32
33
34

35 13 **Figure 4. Structural and topological comparison of thus-far available betaCoV RBD**
36
37 14 **structures.** Four structures, including those of HKU9-, SARS-, MERS- and HKU4-RBD, were
38
39 15 oriented similarly and presented as cartoons in parallel. The core-center, core-peripheral and the
40
41 16 external subdomain are highlighted by encircling in yellow. For each structure, the topological
42
43 17 arrangement of the core-center and core-peripheral strands as well as of the external components
44
45 18 is depicted. The core-strands that flank the external subdomain are colored red and blue,
46
47 19 respectively. Yellow lines indicate disulphide bonds. The N- and C-terminus are highlighted. (A)
48
49 20 HKU9-RBD. (B) SARS-RBD. (C) MERS-RBD. (D) HKU4-RBD. The similarity in the
50
51 21 topological arrangement of the external subdomain as an insertion between two core-strands was
52
53 22 illustrated.
54
55
56
57
58
59
60

Figure 5. Homologous inter-subdomain amino acid interactions anchoring the external subdomain to the core subdomain. (A) Superimposition of the betaCoV RBD (HKU9-RBD in green, SARS-RBD in yellow, MERS-RBD in blue and HKU4-RBD in cyan) structures highlighting the external elements that can be well-aligned. These two elements, with 7 (element 1) and 8 (element 2) amino acids, respectively, engage mainly the core subdomain H2 and H6 helices for the inter-subdomain interactions. To facilitate comparison, the element residues were successively assigned a position marker (a-g for element 1 and a-h for element 2), which is highlighted. (B) Characterization of the element residues for their contributions to the inter-subdomain binding. The two external elements were presented as cartoons, while the core subdomain is shown in surface. At each position, the residue is marked sequentially with the position marker, the amino acid identity and numbering, the interacting mode/type and the side-chain orientation. For the interaction mode, the hydrophobic or van der Waals interactions are indicated with encircled Ps, the side-chain H-bonds with encircled Ss and main-chain H-bonds with encircled Ms. The side-chain orientations are indicated with arrows. (C) Summarization of the inter-subdomain interactions specified in (B). The element sequences of the four RBDs were aligned and listed. “+” indicates that a certain type of interaction is commonly observed at the position, while “+/-” indicates that the interaction type is specific to some but not all of the four RBDs. The arrows mark the side-chain orientations.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 TABLES

2 **Table 1. Data collection and refinement statistics.**

Data collection	HKU9-RBD (PDB code 5GYQ)	Au derivative HKU9-RBD
Space group	P21	P1
Wavelength	1.03906	1.03906
Unit cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	42.7, 36.0, 62.9	36.0, 46.6, 57.3
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90.0, 102.7, 90.0	80.4, 88.8, 88.5
Resolution ^a (Å)	50.00-2.10 (2.18-2.10)	50.00-2.48 (2.57-2.48)
Observed reflections	101,588	52,401
Completeness (%)	97.1 (80.7)	97.7 (96.9)
Redundancy	9.4 (9.4)	4.1 (3.7)
<i>R</i> _{merge} ^b (%)	6.1 (15.4)	9.2 (39.0)
<i>I</i> / <i>σI</i>	34.023 (12.057)	16.960 (4.637)
<i>CC</i> _{1/2}	0.998 (0.988)	0.986 (0.915)
Refinement		
Resolution (Å)	41.7-2.10	
Number of reflections	10,811	
Completeness for range (%)	97.0	
<i>R</i> _{work} / <i>R</i> _{free} ^c	0.1700/0.2006	
No. atoms		
Protein	1367	
Water	128	
B-factors		
Protein	28.7	

Water	34.1
R.m.s. deviations	
Bond length (Å)	0.003
Bond angles (°)	0.820
Ramachandranplot ^d	
Favored (%)	94.64
Allowed (%)	5.36
Outliers (%)	0.00

^a Values for the outmost resolution shell are given in parentheses.

^b $R_{\text{merge}} = \frac{\sum_i \sum_{hkl} |I_i - \langle I \rangle|}{\sum_i \sum_{hkl} I_i}$, where I_i is the observed intensity and $\langle I \rangle$ is the average intensity from multiple measurements.

^c $R_{\text{work}} = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$, where F_o and F_c are the structure-factor amplitudes from the data and the model, respectively. R_{free} is the R factor for a subset (5%) of reflections that was selected prior to refinement calculations and was not included in the refinement.

^d Ramachandran plots were generated by using the program MolProbity.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2. Statistics of the core subdomain deviations among thus-far available betaCoV RBD structures.

	HKU9-RBD	MERS-RBD	SARS-RBD	HKU4-RBD
HKU9-RBD	—	2.07 Å (104 Cα)	1.92 Å (100 Cα)	1.37 Å (94 Cα)
MERS-RBD		—	2.82 Å (75 Cα)	0.66 Å (109 Cα)
SARS-RBD			—	1.95 Å (82 Cα)
HKU4-RBD				—

The RBD core subdomain structures were superimposed onto each other by PyMol in a pairwise manner to calculate the root-mean-square deviation (rmsd) values, which were listed in the table. The values in parentheses indicate the number of equivalent Cα atoms that were selected for rmsd calculations.

AUTHOR INFORMATION

Corresponding Author

*Tel: 86-10-64807688; Fax: 86-10-64807882; Email: gaof@im.ac.cn, gaofu@chinacdc.cn.

Funding Sources

This work was supported by the National Natural Science Foundation of China (NSFC, Grant No. 81461168030 and 81290342), the China National Grand S&T Special Project (Grant No. 2014ZX10004-001-006), the Strategic Priority Research Program of the Chinese Academy of

Sciences Grant XDB08020100. G. L. is supported in part by NSFC (Grant No. 31400154, 81522026 and 31570157) and the Sichuan Outstanding Youth Science & Technology Funding (Grant No: 2016JQ0001). G.F.G. is a leading principle investigator of the NSFC Innovative Research Group (Grant No. 81321063).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Assistance in diffraction data collection by the staff at the Shanghai Synchrotron Radiation Facility (SSRF-beamline 17U) is acknowledged. We thank Dr. Zheng Fan (Institute of Microbiology, Chinese Academy of Sciences) and Yuanyuan Chen, Zhenwei Yang (Institute of Biophysics, Chinese Academy of Sciences) for their sophisticated technical supports in the SPR assay.

REFERENCES

- [1] Fields, B. N., Knipe, D. M., and Howley, P. M. (2007) *Fields virology*, 5th ed., Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.
- [2] Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C., Zhou, J., Liu, W., Bi, Y., and Gao, G. F. (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses, *Trends Microbiol* 24, 490-502.
- [3] International Committee on Taxonomy of Viruses., and King, A. M. Q. (2012) *Virus taxonomy : classification and nomenclature of viruses : ninth report of the International Committee on Taxonomy of Viruses*, Academic Press, London ; Waltham, MA.
- [4] Schalk, A. F., and Hawn, M. C. (1931) An apparently new respiratory disease in baby chicks., *Journal of the American Veterinary Medical Association* 78, 413-422.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

[5] Woo, P. C., Lau, S. K., Lam, C. S., Lai, K. K., Huang, Y., Lee, P., Luk, G. S., Dyrting, K. C., Chan, K. H., and Yuen, K. Y. (2009) Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus, *Journal of virology* 83, 908-917.

[6] Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., Bai, R., Teng, J. L., Tsang, C. C., Wang, M., Zheng, B. J., Chan, K. H., and Yuen, K. Y. (2012) Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus, *Journal of virology* 86, 3995-4008.

[7] Weinstein, R. A. (2004) Planning for epidemics--the lessons of SARS, *The New England journal of medicine* 350, 2332-2334.

[8] Organization, W. H. (2014) Global Alert and Response (GAR): MERS-CoV summary updates, http://www.who.int/csr/disease/coronavirus_infections/archive_updates/en/.

[9] Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., and Wang, L. F. (2005) Bats are natural reservoirs of SARS-like coronaviruses, *Science* 310, 676-679.

[10] Woo, P. C., Lau, S. K., Li, K. S., Poon, R. W., Wong, B. H., Tsoi, H. W., Yip, B. C., Huang, Y., Chan, K. H., and Yuen, K. Y. (2006) Molecular diversity of coronaviruses in bats, *Virology* 351, 180-187.

[11] Woo, P. C., Wang, M., Lau, S. K., Xu, H., Poon, R. W., Guo, R., Wong, B. H., Gao, K., Tsoi, H. W., Huang, Y., Li, K. S., Lam, C. S., Chan, K. H., Zheng, B. J., and Yuen, K. Y. (2007) Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features, *Journal of virology* 81, 1574-1585.

[12] Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G., Peiris, M., Lim, W., Stohr, K., and Osterhaus, A. D. (2003) Aetiology: Koch's postulates fulfilled for SARS virus, *Nature* 423, 240.

[13] Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J., Anderson, L. J., and Group, S. W. (2003) A novel coronavirus associated with severe acute respiratory syndrome, *The New England journal of medicine* 348, 1953-1966.

[14] Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H., and Doerr, H. W. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *The New England journal of medicine* 348, 1967-1976.

[15] Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A. D.,

- 1 Drosten, C., Pallansch, M. A., Anderson, L. J., and Bellini, W. J. (2003) Characterization
2 of a novel coronavirus associated with severe acute respiratory syndrome, *Science* 300,
3 1394-1399.
- 4 [16] de Groot, R. J., Baker, S. C., Baric, R. S., Brown, C. S., Drosten, C., Enjuanes, L., Fouchier,
5 R. A., Galiano, M., Gorbalenya, A. E., Memish, Z. A., Perlman, S., Poon, L. L., Snijder,
6 E. J., Stephens, G. M., Woo, P. C., Zaki, A. M., Zambon, M., and Ziebuhr, J. (2013)
7 Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the
8 Coronavirus Study Group, *Journal of virology* 87, 7790-7792.
- 9 [17] Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A.
10 (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, *The*
11 *New England journal of medicine* 367, 1814-1820.
- 12 [18] Bermingham, A., Chand, M. A., Brown, C. S., Aarons, E., Tong, C., Langrish, C., Hoschler,
13 K., Brown, K., Galiano, M., Myers, R., Pebody, R. G., Green, H. K., Boddington, N. L.,
14 Gopal, R., Price, N., Newsholme, W., Drosten, C., Fouchier, R. A., and Zambon, M.
15 (2012) Severe respiratory illness caused by a novel coronavirus, in a patient transferred to
16 the United Kingdom from the Middle East, September 2012, *Euro surveillance : bulletin*
17 *Europeen sur les maladies transmissibles = European communicable disease bulletin* 17,
18 20290.
- 19 [19] Woo, P. C., Lau, S. K., Chu, C. M., Chan, K. H., Tsoi, H. W., Huang, Y., Wong, B. H., Poon,
20 R. W., Cai, J. J., Luk, W. K., Poon, L. L., Wong, S. S., Guan, Y., Peiris, J. S., and Yuen, K.
21 Y. (2005) Characterization and complete genome sequence of a novel coronavirus,
22 coronavirus HKU1, from patients with pneumonia, *Journal of virology* 79, 884-895.
- 23 [20] Lu, G., and Liu, D. (2012) SARS-like virus in the Middle East: a truly bat-related
24 coronavirus causing human diseases, *Protein & cell* 3, 803-805.
- 25 [21] Lu, G., Wang, Q., and Gao, G. F. (2015) Bat-to-human: spike features determining 'host
26 jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond, *Trends in microbiology* 23,
27 468-478.
- 28 [22] Ge, X. Y., Li, J. L., Yang, X. L., Chmura, A. A., Zhu, G., Epstein, J. H., Mazet, J. K., Hu, B.,
29 Zhang, W., Peng, C., Zhang, Y. J., Luo, C. M., Tan, B., Wang, N., Zhu, Y., Crameri, G.,
30 Zhang, S. Y., Wang, L. F., Daszak, P., and Shi, Z. L. (2013) Isolation and characterization
31 of a bat SARS-like coronavirus that uses the ACE2 receptor, *Nature* 503, 535-538.
- 32 [23] Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S.
33 Y., Chan, K. H., and Yuen, K. Y. (2005) Severe acute respiratory syndrome coronavirus-
34 like virus in Chinese horseshoe bats, *Proceedings of the National Academy of Sciences of*
35 *the United States of America* 102, 14040-14045.
- 36 [24] Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. Y., Zheng, K., Lei, L. C., Chen, Q. X., Gao,
37 Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan,
38 H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F., Qin, P. Z., Li, L. H., Ren, Y.
39 Q., Liang, W. J., Yu, Y. D., Anderson, L., Wang, M., Xu, R. H., Wu, X. W., Zheng, H. Y.,
40 Chen, J. D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L. Y., Li, H., Chen, F., Di, B.,
41 He, L. J., Lin, J. Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X. J.,
42 Spiga, O., Guo, Z. M., Pan, H. Y., He, W. Z., Manuguerra, J. C., Fontanet, A., Danchin,
43 A., Niccolai, N., Li, Y. X., Wu, C. I., and Zhao, G. P. (2005) Cross-host evolution of
44 severe acute respiratory syndrome coronavirus in palm civet and human, *Proceedings of*
45 *the National Academy of Sciences of the United States of America* 102, 2430-2435.
- 46 [25] Lau, S. K., Li, K. S., Tsang, A. K., Lam, C. S., Ahmed, S., Chen, H., Chan, K. H., Woo, P.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

C., and Yuen, K. Y. (2013) Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus, *Journal of virology* 87, 8638-8650.

[26] Ithete, N. L., Stoffberg, S., Corman, V. M., Cottontail, V. M., Richards, L. R., Schoeman, M. C., Drosten, C., Drexler, J. F., and Preiser, W. (2013) Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa, *Emerging infectious diseases* 19, 1697-1699.

[27] Corman, V. M., Ithete, N. L., Richards, L. R., Schoeman, M. C., Preiser, W., Drosten, C., and Drexler, J. F. (2014) Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an african bat, *Journal of virology* 88, 11297-11303.

[28] Yang, L., Wu, Z., Ren, X., Yang, F., Zhang, J., He, G., Dong, J., Sun, L., Zhu, Y., Zhang, S., and Jin, Q. (2014) MERS-related betacoronavirus in Vespertilio superans bats, China, *Emerging infectious diseases* 20, 1260-1262.

[29] Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., Alhakeem, R., Durosinsoun, A., Al Asmari, M., Islam, A., Kapoor, A., Briesse, T., Daszak, P., Al Rabeeah, A. A., and Lipkin, W. I. (2013) Middle East respiratory syndrome coronavirus in bats, Saudi Arabia, *Emerging infectious diseases* 19, 1819-1823.

[30] De Benedictis, P., Marciano, S., Scaravelli, D., Priori, P., Zecchin, B., Capua, I., Monne, I., and Cattoli, G. (2014) Alpha and lineage C betaCoV infections in Italian bats, *Virus genes* 48, 366-371.

[31] Annan, A., Baldwin, H. J., Corman, V. M., Klose, S. M., Owusu, M., Nkrumah, E. E., Badu, E. K., Anti, P., Agbenyega, O., Meyer, B., Oppong, S., Sarkodie, Y. A., Kalko, E. K., Lina, P. H., Godlevska, E. V., Reusken, C., Seebens, A., Gloza-Rausch, F., Vallo, P., Tschapka, M., Drosten, C., and Drexler, J. F. (2013) Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, *Emerging infectious diseases* 19, 456-459.

[32] Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., and Madani, T. A. (2014) Evidence for camel-to-human transmission of MERS coronavirus, *The New England journal of medicine* 370, 2499-2505.

[33] Reusken, C. B., Haagmans, B. L., Muller, M. A., Gutierrez, C., Godeke, G. J., Meyer, B., Muth, D., Raj, V. S., Smits-De Vries, L., Corman, V. M., Drexler, J. F., Smits, S. L., El Tahir, Y. E., De Sousa, R., van Beek, J., Nowotny, N., van Maanen, K., Hidalgo-Hermoso, E., Bosch, B. J., Rottier, P., Osterhaus, A., Gortazar-Schmidt, C., Drosten, C., and Koopmans, M. P. (2013) Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study, *The Lancet. Infectious diseases* 13, 859-866.

[34] Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H., Muller, M. A., Dijkman, R., Muth, D., Demmers, J. A., Zaki, A., Fouchier, R. A., Thiel, V., Drosten, C., Rottier, P. J., Osterhaus, A. D., Bosch, B. J., and Haagmans, B. L. (2013) Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC, *Nature* 495, 251-254.

[35] Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., Ji, W., Li, Y., Wu, Y., Wang, J., Iwamoto, A., Woo, P. C., Yuen, K. Y., Yan, J., Lu, G., and Gao, G. F. (2014) Bat Origins of MERS-CoV Supported by Bat Coronavirus HKU4 Usage of Human Receptor CD26, *Cell host & microbe* 16, 328-337.

[36] Yang, Y., Du, L., Liu, C., Wang, L., Ma, C., Tang, J., Baric, R. S., Jiang, S., and Li, F. (2014)

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
- Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus, *Proceedings of the National Academy of Sciences of the United States of America* 111, 12516-12521.
- [37] Wang, Q., Wong, G., Lu, G., Yan, J., and Gao, G. F. (2016) MERS-CoV spike protein: Targets for vaccines and therapeutics, *Antiviral Res* 133, 165-177.
- [38] Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor, *Science* 309, 1864-1868.
- [39] Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., Zhang, Y., Zhang, W., Yuan, Y., Bao, J., Zhang, B., Shi, Y., Yan, J., and Gao, G. F. (2013) Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26, *Nature* 500, 227-231.
- [40] Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K. C., Chen, Y. H., Zhang, L., and Wang, X. (2013) Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4, *Cell Res* 23, 986-993.
- [41] Peng, G., Sun, D., Rajashankar, K. R., Qian, Z., Holmes, K. V., and Li, F. (2011) Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor, *Proceedings of the National Academy of Sciences of the United States of America* 108, 10696-10701.
- [42] Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature* 426, 450-454.
- [43] Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., and Shi, Z. (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China, *Journal of virology* 86, 4620-4630.
- [44] Lau, S. K., Poon, R. W., Wong, B. H., Wang, M., Huang, Y., Xu, H., Guo, R., Li, K. S., Gao, K., Chan, K. H., Zheng, B. J., Woo, P. C., and Yuen, K. Y. (2010) Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup, *Journal of virology* 84, 11385-11394.
- [45] Tao, Y., Tang, K., Shi, M., Conrardy, C., Li, K. S., Lau, S. K., Anderson, L. J., and Tong, S. (2012) Genomic characterization of seven distinct bat coronaviruses in Kenya, *Virus research* 167, 67-73.
- [46] Tong, S., Conrardy, C., Ruone, S., Kuzmin, I. V., Guo, X., Tao, Y., Niezgoda, M., Haynes, L., Agwanda, B., Breiman, R. F., Anderson, L. J., and Rupprecht, C. E. (2009) Detection of novel SARS-like and other coronaviruses in bats from Kenya, *Emerging infectious diseases* 15, 482-485.
- [47] Zhang, W., Qi, J., Shi, Y., Li, Q., Gao, F., Sun, Y., Lu, X., Lu, Q., Vavricka, C. J., Liu, D., Yan, J., and Gao, G. F. (2010) Crystal structure of the swine-origin A (H1N1)-2009 influenza A virus hemagglutinin (HA) reveals similar antigenicity to that of the 1918 pandemic virus, *Protein & cell* 1, 459-467.
- [48] Song, H., Qi, J., Khedri, Z., Diaz, S., Yu, H., Chen, X., Varki, A., Shi, Y., and Gao, G. F. (2016) An open receptor-binding cavity of Hemagglutinin-Esterase-Fusion glycoprotein from newly-identified influenza D virus: basis for its broad cell tropism, *PLoS Pathog* 12, e1005411.
- [49] Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data, *Methods*

1
2
3 1 *enzymol* 276, 307-326.
4 2 [50] Uson, I., and Sheldrick, G. M. (1999) Advances in direct methods for protein
5 3 crystallography, *Current opinion in structural biology* 9, 643-648.
6 4 [51] Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum
7 5 likelihood, *Acta crystallographica. Section D, Biological crystallography* 57, 1373-1382.
8 6 [52] Cowtan, K. D., and Zhang, K. Y. (1999) Density modification for macromolecular phase
9 7 improvement, *Progress in biophysics and molecular biology* 72, 245-270.
10 8 [53] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J.
11 9 J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W.,
12 10 Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and
13 11 Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular
14 12 structure solution, *Acta crystallographica. Section D, Biological crystallography* 66, 213-
15 13 221.
16 14 [54] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta*
17 15 *crystallographica. Section D, Biological crystallography* 60, 2126-2132.
18 16 [55] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G.
19 17 J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom
20 18 structure validation for macromolecular crystallography, *Acta crystallographica. Section*
21 19 *D, Biological crystallography* 66, 12-21.
22 20 [56] Chen, Y., Rajashankar, K. R., Yang, Y., Agnihothram, S. S., Liu, C., Lin, Y. L., Baric, R. S.,
23 21 and Li, F. (2013) Crystal structure of the receptor-binding domain from newly emerged
24 22 Middle East respiratory syndrome coronavirus, *Journal of virology* 87, 10777-10783.
25 23 [57] Shi, Z. (2010) Bat and virus, *Protein & cell* 1, 109-114.
26 24 [58] Kubo, H., Yamada, Y. K., and Taguchi, F. (1994) Localization of neutralizing epitopes and
27 25 the receptor-binding site within the amino-terminal 330 amino acids of the murine
28 26 coronavirus spike protein, *Journal of virology* 68, 5403-5410.
29 27 [59] Wu, K., Li, W., Peng, G., and Li, F. (2009) Crystal structure of NL63 respiratory
30 28 coronavirus receptor-binding domain complexed with its human receptor, *Proceedings of*
31 29 *the National Academy of Sciences of the United States of America* 106, 19970-19974.
32 30 [60] Reguera, J., Santiago, C., Mudgal, G., Ordone, D., Enjuanes, L., and Casasnovas, J. M.
33 31 (2012) Structural bases of coronavirus attachment to host aminopeptidase N and its
34 32 inhibition by neutralizing antibodies, *PLoS Pathog* 8, e1002859.
35 33
36
37
38
39
40
41
42
43
44 34
45
46
47
48 35
49
50
51
52
53
54
55
56
57
58
59
60

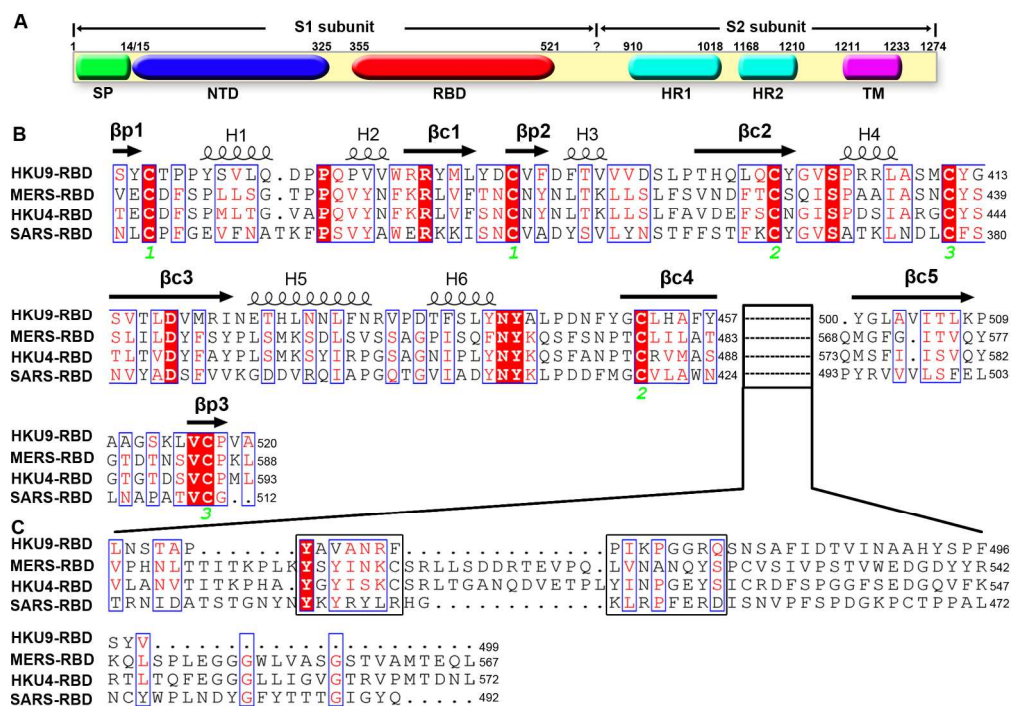


Figure 1. Sequence features of HKU9-RBD

180x127mm (300 x 300 DPI)

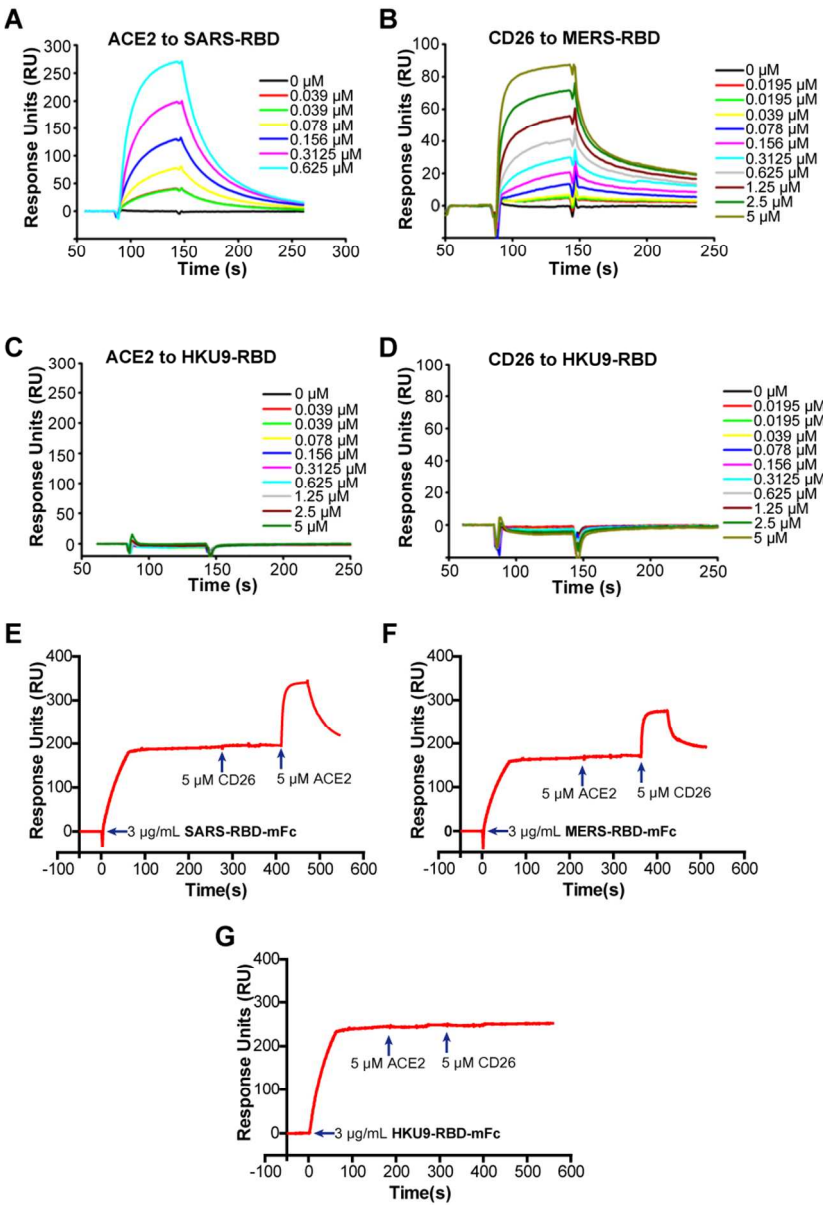


Figure 2. Characterization of HKU9-RBD by SPR assays.

85x123mm (300 x 300 DPI)

85x60mm (300 x 300 DPI)

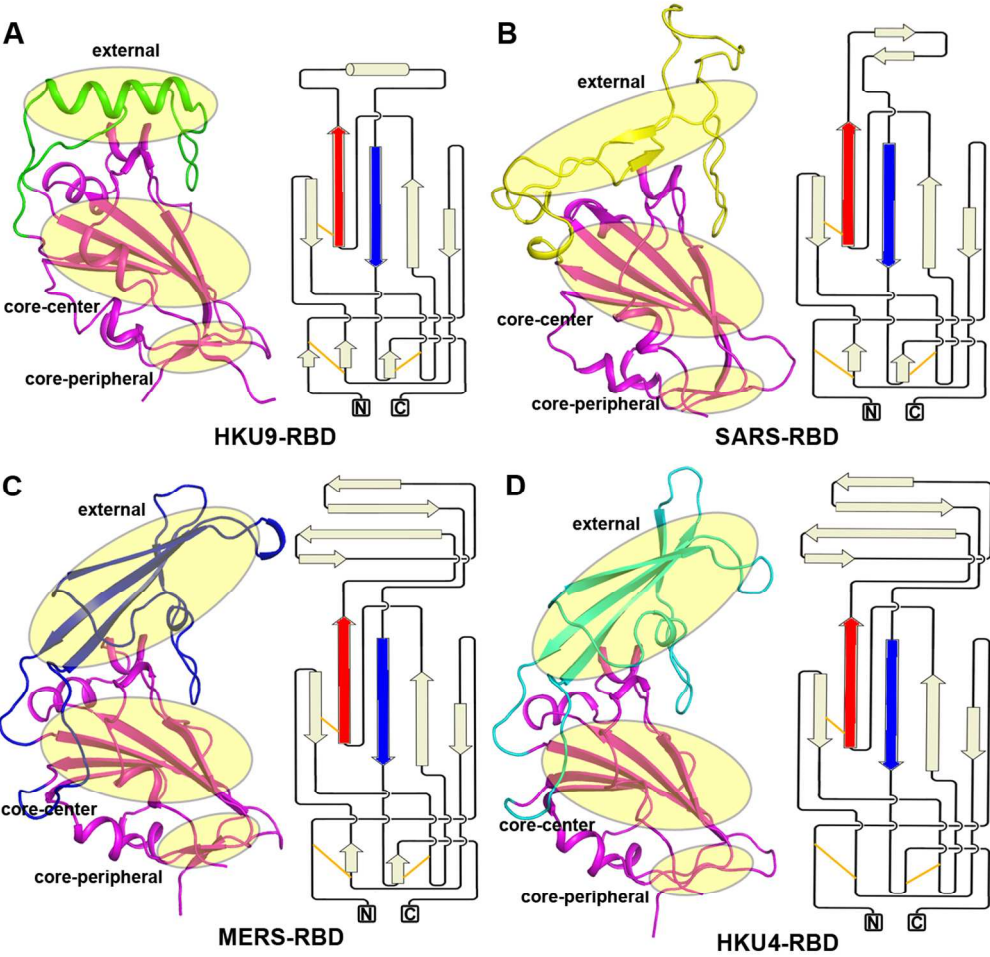


Figure 4. Structural and topological comparison of thus-far available betaCoV RBD structures

119x119mm (300 x 300 DPI)

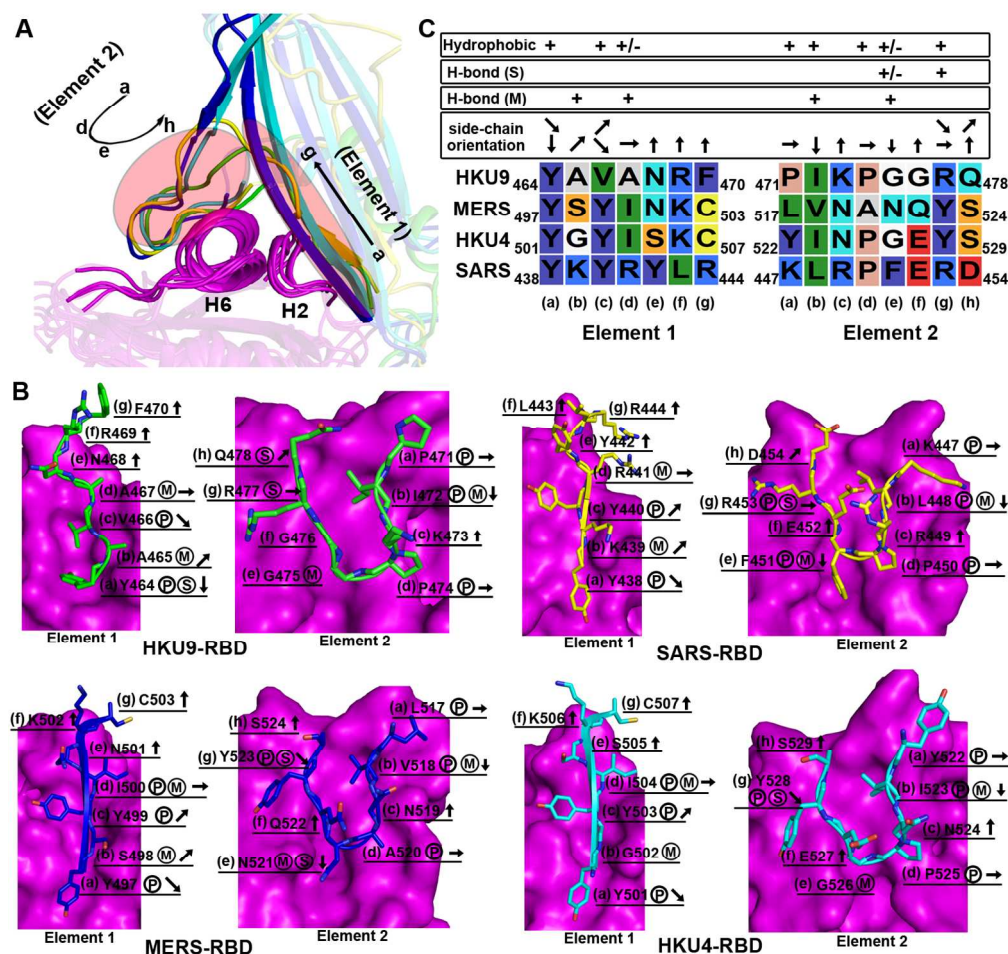


Figure 5. Homologous inter-subdomain amino acid interactions anchoring the external subdomain to the core subdomain

140x134mm (300 x 300 DPI)

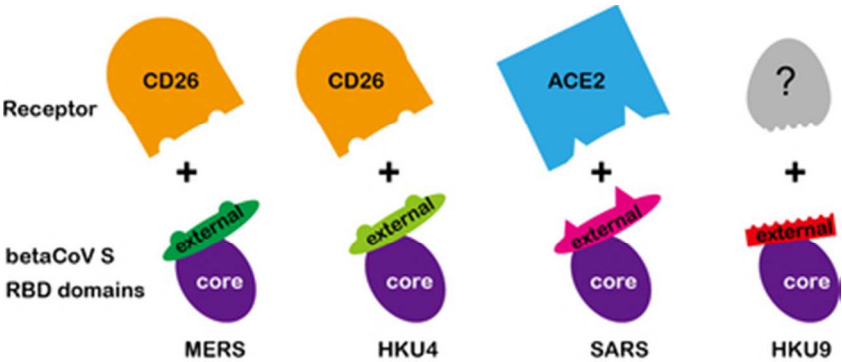


Table of contents (TOC)

35x15mm (300 x 300 DPI)